UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

DTIC FILE COPY

2

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

AD-A192 661

DTIC ELECTE MAR 2 4 1988 S E

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
ULE	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) R-131	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION USAETL 6b. OFFICE SYMBOL (If applicable) CEETL-LO	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Fort Belvoir, VA 22060-5546	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Research for Reducing the Labor Intensive Nature of High-Resolution Terrain Analysis Feature Extraction

12. PERSONAL AUTHOR(S)
DANIEL EDWARDS

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Research	FROM ______ TO ______	87 OCT 8	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	spatial data, data compilation terrain knowledge

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Demands for digital terrain spatial data are rapidly increasing and will continue to grow. Spatial terrain knowledge is critical to the solution of many existing and emerging problems. However, spatial terrain data compilation is a manual, labor-intensive, error-prone process. Requirements for high-resolution terrain data call for innovative approaches to the problem of compilation. Based on terrain analyst productivity estimates of 1000 man-hours per 15 by 15 arc-minute area, the time required to complete a single terrain analysis of the world's land surface exceeds several hundred thousand man years. Another cilemma arises from the way we currently store and use spatial data. Current geographic information system techniques emphasize a "brute-force" search approach to spatial storage, query and analysis. If global high-resolution terrain data were available, the response time for certain "brute-force" data base queries might approach the above time estimates for compilation.

The following research strategies are discussed which address the high-resolution dilemmas. First, terrain feature extraction should be approached from a "minimum compilation, maxium

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL F. DARLENE SEYLER	22b. TELEPHONE (Include Area Code) (202) 355-2647 22c. OFFICE SYMBOL CEETL-LO

DD Form 1473, JUN 86

Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

BLOCK #19 - CONT'D:

analysis" strategy. In other words, map only the key terrain components, and gather additional information by thorough analysis and inferencing from this compiled spatial data. This basic approach parallels techniques used extensively in manual photo-based terrain analysis. Secondly, knowledge needs to be incorporated into all phases of terrain data compilation, storage and analysis. Low-level geometric knowledge of spatial features can be used to organize and group data together that are important at a higher symbolic level of terrain understanding. Similarly, high-level knowledge and models of regional factors such as climate and geomorphology can be used to constrain "brute-force" search, detect errors and handle incomplete information. Exploitation of terrain knowledge in digital spatial information technology can reduce the "data rich" requirement and "knowledge poor" state of current systems. Finally, additional research in the quantification of the symbolic concepts within terrain analysis is needed.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

QUALITY INSPECTED 2

Research for Reducing the Labor Intensive Nature of High-Resolution Terrain Analysis Feature Extraction

Daniel Edwards
U.S. Army Engineer Topographic Laboratories
Topographic Developments Laboratory
Fort Belvoir, VA 22060-5546
(202) 355-2789

BIOGRAPHICAL SKETCH

Daniel Edwards has been a physical scientist with the U.S. Army Engineer Topographic Laboratories (USAETL) since 1981. He served as visiting scientist in the Computer Science Department of Carnegie-Mellon University on the MAPS project during 1986-87. He has worked on computer-assisted photo interpretation, spatial data compilation, storage and analysis techniques. He served as a consultant to the Ohio Department of Natural Resources regarding delineation of surface mines using MSS data. He is a honor graduate of Ohio State University with B.S./M.S. degrees in Natural Resources.

ABSTRACT

Demands for digital terrain spatial data are rapidly increasing and will continue to grow. Spatial terrain knowledge is critical to the solution of many existing and emerging problems. However, spatial terrain data compilation is a manual, labor-intensive, error-prone process. Requirements for high-resolution terrain data call for innovative approaches to the problem of compilation. Based on terrain analyst productivity estimates of 1000 man-hours per 15 by 15 arc-minute area, the time required to complete a single terrain analysis of the world's land surface exceeds several hundred thousand man-years. Another dilemma arises from the way we currently store and use spatial data. Current geographic information system techniques emphasize a "brute-force" search approach to spatial storage, query and analysis. If global high-resolution terrain data were available, the response time for certain "brute-force" data base queries might approach the above time estimates for compilation.

The following research strategies are discussed which address the high-resolution dilemmas. First, terrain feature extraction should be approached from a "minimum compilation, maximum analysis" strategy. In other words, map only the key terrain components, and gather additional information by thorough analysis and inferencing from this compiled spatial data. This basic approach parallels techniques used extensively in manual photo-based terrain analysis. Secondly, knowledge needs to be incorporated into all phases of terrain data compilation, storage and analysis. Low-level geometric knowledge of spatial features can be used to organize and group data together that are important at a higher symbolic level of terrain understanding. Similarly, high-level knowledge and models of regional factors such as climate and geomorphology can be used to constrain "brute-force" search, detect errors and handle incomplete information. Exploitation of terrain knowledge in digital spatial information technology can reduce the "data rich"

88 3 22 053

requirement and "knowledge poor" state of current systems. Finally, additional research in the quantification of the symbolic concepts within terrain analysis is needed.

GENERAL PROBLEM

Seventeenth century philosopher/scientist Francis Bacon wrote that "knowledge and human power are synonymous" (Bacon, 1620). Bacon's maxim helps to explain the current recognition of the importance of spatial information and the increasing demand for it. Requirements, not only for data availability, but also for an exceedingly inordinate breadth and depth of information are emerging (Edwards, 1986). Certainly, solutions to many existing problems and applications either require or would be enhanced by the addition of spatial terrain knowledge. However, spatial terrain data compilation is a manual, labor-intensive, error-prone process. Based on terrain analyst productivity estimates of 1000 man-hours to map a 1:50,000 scale 15 arc-minute by 15 arc-minute area, the time required to complete a single terrain analysis of the world's land surface exceeds several hundred thousand man-years. Ironically, current data base information retrieval techniques are based on the fragile assumption that the specific data sought has been previously stored. The glaring deficiency of this dilemma is that terrain information is compiled, stored and accessed separately with little recognition of the rich interconnections between the data or the inferences that can be made from such data. Given the norm, where existing digital spatial information is incomplete and sometimes inaccurate, the exploitation of known terrain relationships can enhance feature/attribute compilation and provide a means of handling incompleteness and uncertainty.

PARTIAL SOLUTION

There is an area of promise which has been overshadowed by the digital evolution of the past several decades and largely neglected by the mapping community. A more conservative, economical approach to spatial data compilation requires careful mapping of the important components of the terrain and then a wiser, more comprehensive exploitation of that which has been collected. This approach is based on the principles of photo-based terrain analysis originally stated by Frost (Frost,1953):

1) An air photo is a pictorial representation of the various features within the landscape and is composed of pattern elements that serve as indicators of materials, conditions and events that are related to the physical, biological, cultural and climatic components of the landscape.

2) Similar materials and conditions in similar environments produce similar patterns, and unlike materials and conditions produce unlike patterns.

Terrain analysis is based on extraction of key terrain features, i.e. pattern elements, which are then analyzed individually and

collectively in order to form a complete representation of the terrain. The pattern elements mentioned above in (1) consist of : topographic shape, surface drainage pattern and density, drainage gully profile/gradient, vegetation, land use, unique features and photo tone. Analysis of the pattern elements is performed both separately and jointly, based on principle (2) above. Relationships are studied, inferences are made, and conclusions are drawn with the sole purpose of understanding the structure and organization of a particular terrain, so that sufficient information is gathered to permit informed decision-making. It is the values of the pattern elements and their association with one another that defines a particular landform, such as sandstone hill, limestone valley, outwash plain, or volcanoe. In terrain analysis, the mapping of the features present in a given area is only the first step of information compilation. Conversely, in current data base technology, feature mapping is usually the sole compilation step.

By exploiting both the low-level feature and high-level regional knowledge commonly used during terrain analysis and incorporating this in a spatial data base, the amount of storable information can be greatly expanded, errors within the data may be detected, and functional capability may be preserved when information is incomplete. For instance, low-level information such as feature coordinate sets are an existing component of any spatial data base, but are rarely exploited. Such coordinate information could be used to calculate and store geometric descriptors or relationships (linearity, azimuth, slope, shape, rank, etc.). This type of information is vital to organizing and grouping data according to membership of some important symbolic group, such as drainage pattern, or landform type. By such exploitation and analysis, it would be possible to greatly expand the existing knowledge in the data base. Similarly, high-level knowledge such as global climate, geology or vegetative communities is non-existent in present spatial information systems. Through incorporation and exploitation of high-level understanding and models, errors in the data could be detected, query search could be made more efficient, and incomplete information could be accommodated.

REVISING GEOGRAPHIC INFORMATION SYSTEM ASSUMPTIONS

Current GIS technology can be divided into three major components:

1) Data Extraction and Feature Compilation

2) Data Storage and Organization

3) Data Query, Analysis and Visualization

These components are commonly considered as sequentially ordered, separate functions. Logan and Bryant (Logan, 1987) noted that "data typically flows only one way, from digitizer to the GIS system." The above three-tiered GIS model is based on the concept that the first component will be able to supply the data required and that the data are correct. However, due to the arduous nature of feature extraction from source materials, the fundamental assumptions of data availability

and correctness/accuracy are too often violated. Because of this impasse, the above GIS model must be restructured at and between each level.

First, data extraction must be modified to include the concept of feature compilation as an iterative analysis and extraction process based on a min/max strategy. Actual feature extraction must be kept to a minimum. Analysis of mapped features combined with regional/global knowledge should be performed to its fullest extent to maximize data extraction through statistical analysis, mathematical models, knowledge-based models and inferencing. This methodology applies not only to manually digitized features, but also to features extracted through digital image processing or computer vision techniques. This fundamental change in the feature extraction philosophy requires changes in the other two GIS components.

Data organization and storage must be flexible and dynamic. Data structures and architectures must:
a) support flexible hierarchy based on aggregates determined by conceptual and quantitative set membership
b) provide bi-directional parent/child and inheritance links
c) permit storage of quantitative, symbolic and temporary properties at all hierarchical levels
For example, a particular stream segment could be a member of the following successive hierarchy of aggregates: dendritic drainage pattern, all streams, Occoquan River Basin, Potomac River Basin, etc. Bi-directional links would be important for access. Storage of multiple attributes would be key to the grouping and organization of aggregates.

Data query, analysis and exploitation capabilities must:
a) provide access to high-level knowledge such as regional/global knowledge through parent/child/inheritance or spatial computations if there is a great separation in the hierarchy.
b) provide fully implemented boolean logic query capabilities
c) permit use of rule-based mathematical and conceptual models
d) provide access to symbolic and quantitative computation techniques in order to permit symbolic-to-quantitative translations.
Useful high-level information would include global climate, physiography, soil, flora and crops. Access would be through links or point-in-polygon computations against their spatial boundaries.

Finally, this tiered GIS model can no longer be sequentially ordered and limited in function. Logan and Bryant (Logan,1987) stated the need for bi-directional flow of data. This modularity and flexibility in GIS tools are the key to the min/max strategy of iterative mapping and reasoning. For instance, some of the analysis and advanced graphics tools could be indispensable for assisting the compilation process. Likewise, some compilation capabilities are required for storing key conceptual entities or aggregates gathered during the analysis process. At some point, these basic separate components need to be merged into a combined capability so that one can perform data capture, storage/access, query, analysis and portrayal from any level.

NEEDED RESEARCH

As suggested in the above discussion, research in three distinct, but interelated areas is required. An additional consideration is that the discipline and philosophy of terrain analysis is not easily translatable to a digital world. These following investigative areas will become increasingly inter-twined as work progresses. These are:

1. research in terrain analysis-based expert systems which are dedicated to perform a particular aspect of terrain analysis, such as landform classification or drainage network analysis. Initially, some expert systems would be based on largely symbolic data input by an analyst. More advanced implementations would increasingly rely on direct analysis of stored spatial data.

2. research in data base strategies and organizations that will permit storage and access of compiled/computed feature properties, inferred properties/relations, data aggregated into groups by select properties/relationships, and bi-directional hierarchies of spatial information and knowledge.

3. research and development of quantitative descriptions and relationships which define the symbolic descriptors used in the terrain analysis expert systems. For example, quantitative descriptors are required that encompass such qualitative identifiers as "dense," "steep," "hummocky," "dark," "dendritic," "mountainous," "parallel," "v-shape," "mottled," and so on. These quantitative descriptors would be used to interface the expert systems to the terrain analysis data base (thus, reducing the human interface in the first research area). These expert systems would be used both to expand the knowledge in the data base and detect possible qualitative/quantitative errors.

EXPERT SYSTEM RESEARCH

To date, ongoing research has focused primarily on the first two of the three above areas. Development of an expert system devoted to landform classification, entitled TOPOGRAPHER, has been initiated. Geomorphologic analysis provides a hierarchical organization of terrain data from which a limited amount of data can be exploited, inferences can be made, and errors detected. The project objective is to produce a system which gathers basic pattern element information, makes inferences from this input data, concludes the proper geomorphological landform, and makes further inferences. The long-range system objective is to establish a method of detecting errors in basic qualitative terrain information and greatly expand this terrain knowledge through supporting inferences. This system is implemented in OPS5 (Forgy,1981) and has the general framework as diagramed in Figure 1.

TOPOGRAPHER operates on basic pattern element information gathered from the photo analyst (Frost,1953). Next, inferences are made based on this input data. Any errors in the input data are detected through conflicts in supporting inferences and then corrected. The validated input data and inferences are collected into a frame. This frame is then matched against classical landform types. High-level regional

TOPOGRAPHER STRATEGY

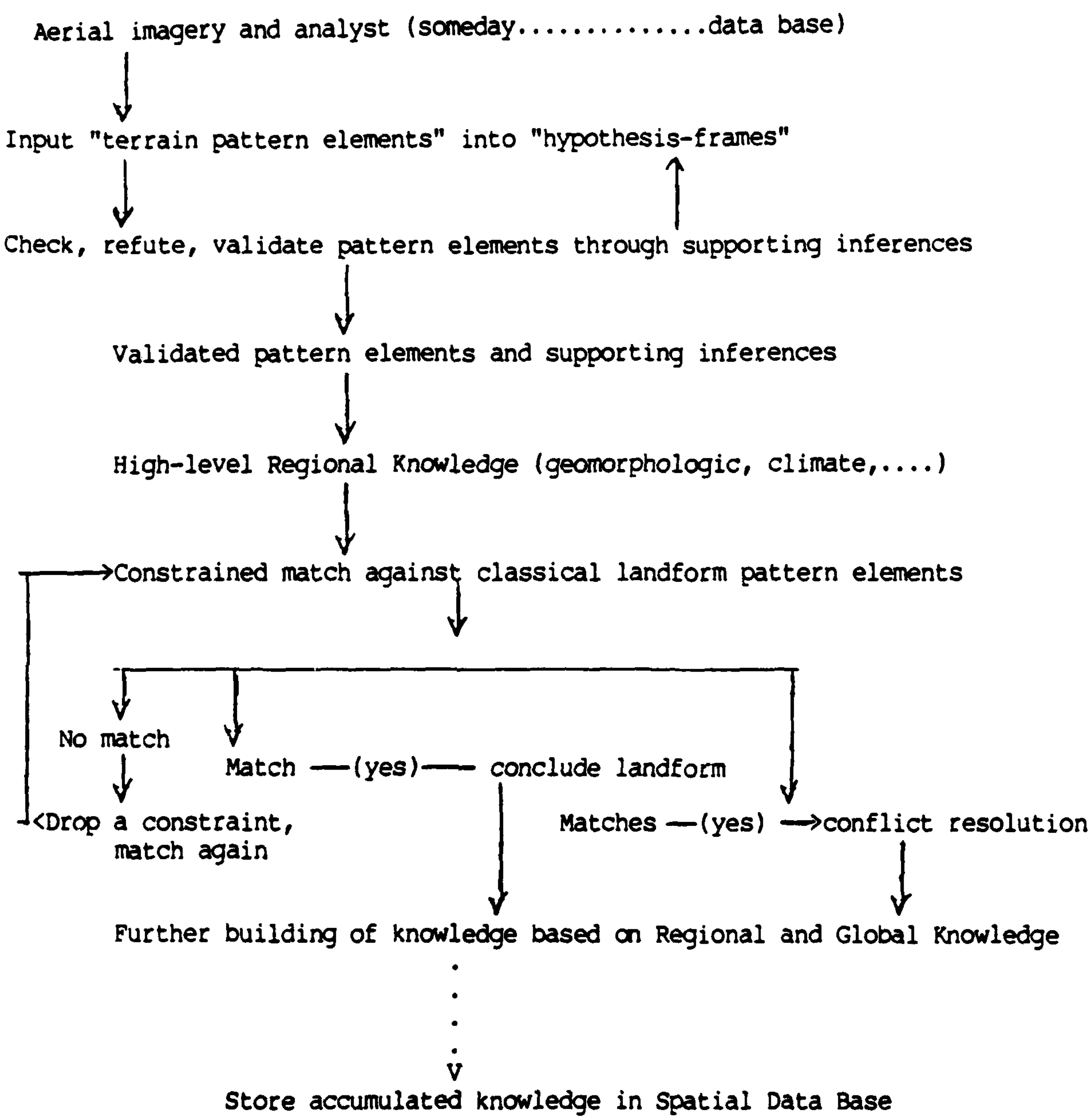

Figure 1.

geomorphic and climate knowledge is used to reduce search. Single matches against classical landforms are displayed to the analyst as the probable landform. Multiple matches are resolved through specific landform conflict resolution rules (which are not yet implemented). If no match occurs, a constraint in the input frame is removed, and then this modified frame is matched against the classical landform pattern elements again. Once a landform is concluded, further knowledge is gathered through inferences, computations and other regional knowledge for the sole purpose of accumulating and storing information.

DATA BASE RESEARCH

Research in data base organizations that would support this iterative compilation/analysis method for feature extraction has been based on the ongoing MAPS research project at Carnegie-Mellon University. A linked hierarchical organization is required to support the reasoning process and information levels required by terrain-based expert systems. This organization is also vital to the query and analysis process. Research, to date, has been based on the MAPS data base (McKeown, 1987) and has been used to partially implement the following levels which are organized through a conceptual hierarchy, and also accessible through a computable spatial hierarchy (McKeown, 1984). The following levels of information are not necessarily fixed in hierarchical order and may not be needed in all applications.

Entity

The lowest hierarchical level is the individual feature/entity level where feature boundary coordinates and attributes are stored. In addition, spatially descriptive measures of the feature (length, width, slope, area, linearity,rank..) should also be stored. These are usually, but not necessarily, based on calculations made on the feature coordinate set.

Conceptual Aggregate

The next higher level consists of conceptual aggregates. These are collections of features with similar attribute characteristics that define a meaningful symbolic conceptual group that match our current understanding of natural relationships. Common examples would include different drainage patterns, soil groupings, and vegetative classifications. Membership at this level is bound by a set of criteria that defines this conceptual grouping. The descriptive measures stored at this level should be collective summary statistics calculated from its individual members, such as average and standard deviation. This aggregate may or may not be spatially significant, storage of its boundary is optional.

Entity Aggregate

The next level consists of entity aggregates where individual features are grouped by entity descriptions/categories, such as all gullies, buildings or forests. The descriptive measures stored should be a meaningful collective statistic based on its individual members such as mode(s) or range. The spatial extent is probably not very meaningful and calculation is optional.

Inferred Conceptual Aggregate

Other high levels consist of inferred conceptual aggregates. These levels in the hierarchy would define conceptually significant groups where little or no underlying information is present. These are aggregates with no actual members. The group and attribute set is formed through inferences made from other entities/aggregates in the data base. An example at this level would be a class of gravelly soil materials based on drainage pattern/density, gully, and climate inferences. Here information that is needed, but not available, can be inferred, stored and accessed. The method of derivation of such information must be stored, as well as any meaningful statistics that can be inferred or calculated.

Inferred Entity Aggregate

Another set of levels consists of inferred entity aggregates. These define groups of entities where little or no underlying information is present. These are aggregates with no actual members. The group and attribute set is formed through inferences made from other entities/aggregates in the data base. Common examples would be particular vegetation species based on inferences of soil moisture, species requirements, and regional knowledge. Here important information, that is not feasible to map, can be inferred, stored and accessed. The method of derivation of such information must be stored, as well as any meaningful statistics that can be inferred or calculated.

Inferred Entity

Another level consists of inferred entities which define a significant entity where little or no underlying information is present. An entity was formed from a strong or supporting inferences, such as particular landform. The attribute set is formed through inferences made from other entities/aggregates in the data base. Here information that is needed, but not readily mapped, can be inferred, stored and accessed. The method of derivation of such information must be stored, as well as any meaningful statistics that can be inferred, inherited or calculated.

Project/Regional/Global Levels

Finally, additional levels consisting of project(s)/area-of-interest might be stored. Additionally, significant global and regional levels would be stored. This would consist of small-scale information that would be accessible through spatial computation or through links if the distance was not so great in the hierarchy. This would include regionally known information that could be detected through spatial operations and inherited by lower levels, such as global/regional maps of geomorphology, climate, soils, natural vegetation communities, crop types and natural fauna. A general organization of these levels is in Figure 2.

SUMMARY

The goal of the above research is to address the labor intensive nature of feature extraction. The primary goals of TOPOGRAPHER are to greatly expand the knowledge base through supporting inferences and to detect errors in the data. A secondary goal is to conclude the proper landform. Initially, this system requires symbolic data input by an

LINKED HIERARCHICAL ORGANIZATION OF SPATIAL FEATURE TYPES

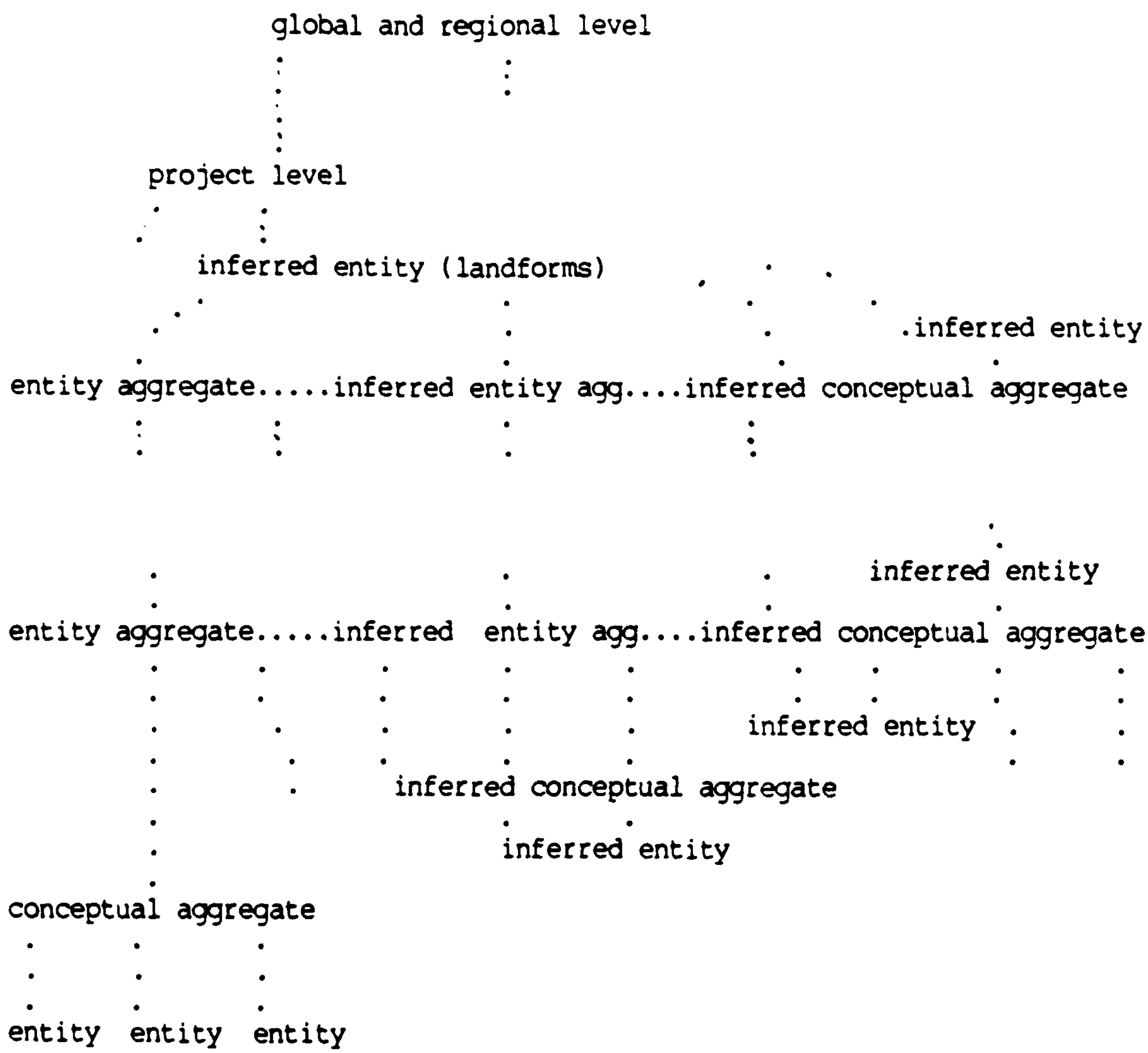

Figure 2

analyst. The ultimate objective is to interface the expert system to a spatial data base through a symbolic-to-quantitative translation. A linked hierarchical data base organization is required for both data input to an expert system and information output from an expert system.

In summary, most complex problems and decisions faced by man require a variety of diverse, but oft inter-related data at varying levels of abstraction. The dilemma of feature extraction is profound. If spatial data base technology is to continue to grow as a useful problem solving tool, then conceptually organized knowledge of physical relationships must be incorporated with stored spatial entities. Access and analysis must involve both stored and inherited relationships. Compilation, storage, analysis, and visualization capabilities must be connected and intertwined.

Acknowledgements

The author would like to thank Dave McKeown, Robert Lai, Wilson Harvey, and the rest of MAPS Project staff for their insight, support, and assistance on many of the concepts in this paper.

References

Bacon, Francis, 1620. Novum Organum.

Edwards, D., Desmond, G., Schoppmann, M. 1986. "Terrain Database Generation for Autonomous Land Vehicle Navigation," Coauthored, International Society for Photogrammetry and Remote Sensing Commission II Symposium, Baltimore, Maryland, May 1986.

Forgy, Charles, 1981. OPS5 User's Manual. Department of Computer Science, Carnegie-Mellon University, Pittsburgh, PA. CMU-CS-81-135.

Frost,R.E., et.al., 1953. A Manual on the Airphoto Interpretation of Soils and Rocks for Engineering Purposes. School of Civil Engineering and Engineering Mechanics, Purdue University, West Lafayette, Indiana.

Logan, Thomas L. and Bryant, Nevin A. 1987. "Spatial Data Software Integration: Merging CAD/CAM/Mapping with GIS and Image Processing," Photogrammetric Engineering and Remote Sensing, Volume LIII, Number 10, pp.1391-1395.

McKeown, D. M. Jr. 1987. "The Role of Artificial Intelligence in the Integration of Remotely Sensed Data with Geographic Information Systems," IEEE Transactions on Geoscience and Remote Sensing, Volume GE-25, Number 3, May, 1987. pp 330-348.

McKeown,D. M. Jr. 1984. "Knowledge-Based Aerial Photo Interpretation," Photogrammetria, Vol. 39, Journal of the International Society for Photogrammetry and Remote Sensing, pp. 91-123.